MLops Assignment-2

Assignment: Enhancing and Optimizing an MLOps Pipeline

Objective:

The goal of this assignment is to extend the MLOps pipeline you've built during the lab sessions. You will explore strategies to enhance feature engineering, model selection, preprocessing techniques, and pipeline automation. The objective remains to predict the number of bike rentals (target variable: cnt) using the Bike Sharing dataset but with added complexity and optimization.

1. Create at least two new interaction features between numerical variables (e.g., `temp * hum`). Justify your choice of features and explain how they might improve the model's predictive performance.
2. Replace the `OneHotEncoder` with `TargetEncoder` for categorical variables. Evaluate how this change impacts the model's performance compared to one-hot encoding.
3. Train LogisticRegressor:
    a. Using the package,
    b. Write/Train it by **scratch** following the steps of a logistic regressor.

    Compare their performance using metrics like Mean Squared Error (MSE) and R-squared.

4. Integrate MLflow into your pipeline to track experiments and save the screenshot of MLflow.
5. Make a report of your results and pipeline.
6. Keep your report and code available in the next lab. We will teach you how to create a new repository in the main branch and push the files into the repository.